DOI:
10.1021/acs.jpclett.6b00518

Document Version
Peer reviewed version

Coarse Master Equations for Binding Kinetics of Amyloid Peptide Dimers

Cathal T. Leahy[1,2], *Ronan D. Murphy*[1,2], *Gerhard Hummer*[3], *Edina Rosta*[4], *and Nicolae-Viorel Buchete*[1,2,*]

[1] School of Physics, University College Dublin, Belfield, Dublin 4, Ireland
[2] Complex and Adaptive Systems Laboratory, University College Dublin, Belfield, Dublin 4, Ireland
[3] Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Max-von-Laue-Straße 3, D-60438 Frankfurt am Main, Germany
[4] Department of Chemistry, King's College London, London SE1 1DB, United Kingdom

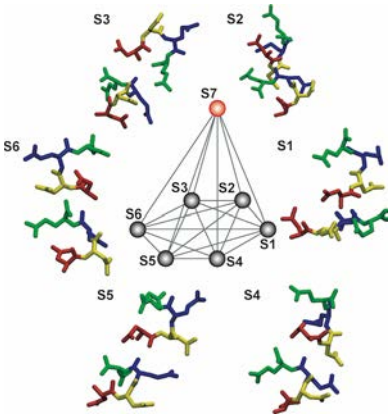
Date: June 17th, 2016

Email: buchete@ucd.ie

ABSTRACT

We characterize the kinetics of dimer formation of the short amyloid microcrystal-forming tetrapeptides NNQQ by constructing coarse master equations for the conformational dynamics of the system, using temperature replica-exchange molecular dynamics (REMD) simulations. We minimize the effects of Kramers-type recrossings by assigning conformational states based on their sequential time evolution. Transition rates are further estimated from short-time state propagators, by maximizing the likelihood that the extracted rates agree with the observed atomistic trajectories without any *a priori* assumptions about their temperature dependence. Here, we evaluate the rates for both continuous replica trajectories that visit different temperatures, and for discontinuous data corresponding to each REMD temperature. While the binding-unbinding kinetic process is clearly Markovian, the conformational dynamics of the bound NNQQ dimer has a complex character. Our kinetic analysis allows us a quantitative discrimination between short-lived encounter pairs and strongly bound conformational states. The conformational dynamics of NNQQ dimers supports a kinetically driven aggregation mechanism, in agreement with the polymorphic character reported for amyloid aggregates such as microcrystals and fibrils.

TOC Graphic



ABBREVIATIONS

REMD = replica-exchange molecular dynamics; MD = molecular dynamics, CME = coarse master equation

Molecular dynamics (MD) simulations of biomolecules play increasingly central roles in complementing a variety of experimental and theoretical studies in fields ranging from material nanoscience to drug design. However, MD studies are continuously challenged by the intrinsic complexity of atomistic systems, and developments well beyond Moore's Law (i.e., hardware improvements) are required to extend their applicability range, particularly to biomolecular systems.[1-6] Modern simulations rely increasingly on advanced enhanced sampling algorithms and analysis methods that help to overcome some of the large data complexity- and size-related limitations of systems such as solvated biomolecules and interacting complexes. [3, 6-7]

Here we show how the complex binding-unbinding dynamics of peptides can be characterized in detail with atomistic MD simulations in an explicit solvent, using an analysis method based on coarse-master equations (CME).[2, 8-9] Without loss of generality, but motivated by computational sampling concerns, we apply our analysis method to the dimerization process of NNQQ peptides, some of the smallest amyloids with biomedical relevance, characterized both theoretically and experimentally in their fibrillar and microcrystalline forms.[10-14] We show how replica-exchange MD (REMD)[9, 15-18] – a powerful and increasingly popular algorithm recently implemented in many atomistic molecular simulation packages – can be used in conjunction with the CME approach to overcome sampling limitations, and to analyze the otherwise complex dynamics of two interacting NNQQ tetrapeptides in explicit water molecules.

Based on possible packing conformations reported for microcrystals, we use four distances, d_1 to d_4, defined as distances between terminal heavy atoms of each side chain (see Fig. 1a). While we have also considered other measures as possible reaction coordinates (RCs), such as the distance between the centers of mass (d_{CM}) or end-to-end distances (d_{EE}), we observe that d_3 and d_4 best separate the population basins. Note that the conformational dynamics of each monomer

could be responsible for different binding modes that would correspond to the same value of a single low-dimensional RC (e.g., d_{CM}). The conformational distributions of our short NNQQ tetrapeptides are rather broad, especially at high temperatures (see d_{EE} histograms in Fig. S1). Here we prefer thus to use the set of two well-defined distances d_3 and d_4 that can capture well the differences in populations between various binding modes (Fig. 1), and, in this case, offer a better discrimination between states then, for example, principal-component based collective variables that offer less local information but can be useful for larger systems.[19-25]

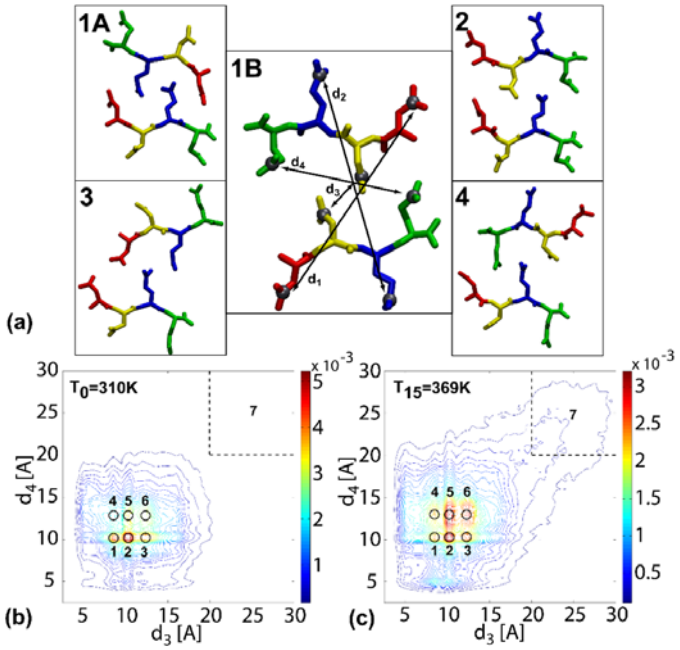


Figure 1. (a) Five initial conditions of the NNQQ dimers. The reaction coordinates are distances d_1, d_2, d_3 and d_4 between the last carbon atom of each residue on one monomer to the corresponding carbon on the other monomer. (b) Contour plot of the projection of d_3 and d_4 showing the normalized populations at T_0 (310 K) and (c) at T_{15} (369 K) used to identify states S_1 to S_7 of the system.

Six population peaks, corresponding to various binding modes denoted by S_1 to S_6, are observed in the population density map in Fig. 1(c) for our highest REMD temperature (T_{15} = 369 K), and for the lowest temperature (T_0 = 310 K) in Fig. 1(b). The dissociated state S_7 is

depicted for a cutoff distance of 20 Å (dashed line). In this study, five microcrystal-like initial conformations (ICs) of NNQQ dimers (see Fig. 1(a)) are used to initialize five independent REMD simulations. The ICs similar to classes 1A, 1B and 4 (see Ref. [11] for the definitions of these structural classes) were taken from the Protein Data Bank (code names 2ONX and 2OLX)[11], with ICs for class 2 and 3 being set up using VMD.[26] The MD package GROMACS[27] was used with the Amber 99sb force field[28] and the dimer is placed in a cubic simulation box of side 40 Å, and is solvated using explicit TIP3P water molecules[29] with periodic boundary conditions. In order to enhance the sampling of the underlying free energy landscape, REMD is used with 16 replicas spanning a temperature range of 310.00 K to 369.08 K with 6526 atoms (130 protein atoms and 2132 water molecules) per replica. The replica temperatures in our REMD runs were optimized (Fig. S6).[30] The results of the CME analysis are not, in principle, dependent on the details of the REMD simulation setup, as long as the data contains sufficient, converged information about the underlying kinetics intrinsic to the system being studied. Each replica is equilibrated at its target temperature with position restraints before running the production MD in the NPT ensemble. We use Langevin dynamics with a friction coefficient of 0.1 ps^{-1},[31] an integration time step of 2 fs, Berendsen pressure coupling,[32] and a particle-mesh Ewald method with switching distance for nonbonded electrostatics and van der Waals interactionsat 8.5 Å and a cutoff distance of 10 Å. Coordinates are saved every 1 ps and REMD exchanges also are attempted every 1 ps, with an average acceptance probability of ~30%. Attempting an exchange as often as possible has been found to enhance the sampling even further.[16, 33] The five ICs are simulated for 160 ns for each replica giving a total REMD running time of 800 ns, and thus a total MD simulation time of 12.8 μs. As shown in Figures S2 and S3

(discussed below), this is more than twice the amount of data needed for convergence of relevant kinetic quantities.

Intermediate states are assigned for the NNQQ dimer using a trajectory-based assignment (TBA) method proven to minimize the effects of fast Kramers-type recrossings, by considering not only instantaneous conformations but also the sequential time evolution of transitions between conformational states.[2, 8] TBA is a simple yet powerful two-step method that allows a direct conversion of multi-dimensional atomistic coordinates to low-dimensional coarse-grained trajectories transitioning between relatively few proposed discrete conformational states. In step 1 of TBA, only the conformational points in the immediate vicinity of population maxima are assigned to their corresponding states (see Fig. 1). In step 2, trajectory points outside these neighborhoods (i.e., unassigned in step 1) are assigned to a certain state if they are located on a transition path that both originates and ends in that state without crossing any other neighborhood boundary. Alternatively, trajectory points on transition paths between different states are assigned to the nearest, already labeled state. As shown previously, this eliminates short, non-reactive Kramers recrossings.[2, 8] Specific to REMD simulations is the fact that the atomistic coordinates are typically saved at each temperature in data files for which we use the term "*T*-trajectories" (i.e., corresponding to one temperature *T*, and thus to all the different replicas evolving at this *T*). To fully characterize the REMD simulation, the history of replica exchange events is also saved separately in corresponding exchange data files. The information from both these two types of files can be used to generate what we denote as "*R*-trajectories" (i.e., corresponding to a replica *R* as it progresses at different temperatures after accepted exchange events).

Importantly, unlike the typical *T*-trajectories, *R*-trajectories are continuous in time and can thus be used to assign states with our TBA method. It is only after this step, by using again the exchange history data, that states can be also assigned accurately along the more typical REMD *T*-trajectories, enabling thus the temperature-dependent investigation of the dynamics.

The state-assigned REMD trajectories, representing transitions between the six bound and one dissociated state (Figs. 1(b) and 1(c)), are further analyzed by collecting short-time propagators (Green's functions), $G(n,\Delta t \mid m,0)$ defined as the conditional probabilities that the system is in state *n* at time Δt, given that it was initially in state *m* at time $t_0 = 0$. The time window Δt is also known as the propagator's "*lag time*". The likelihood function of one continuous Markovian MD trajectory can be written for a system with *N* states[2, 8, 34] as

$$\Lambda_{MD} = \prod_{n,m=0}^{N-1} \left[G(n,\Delta t \mid m,0) \right]^{N_{nm}(\Delta t)} \qquad (1)$$

where $N_{nm}(\Delta t)$is the number of transitions that take a trajectory from state *m* to state *n* after the lag time Δt. Here, we generalize the likelihood function to the two types of trajectories available from an REMD simulation (i.e., *T*- and *R*-trajectories) where we have N_R replicas running at N_T temperatures (commonly, including in our study, $N_R = N_T$). We note that $N_{nm}(\Delta t)$ can be decomposed in values corresponding to a specific replica *R*, with $R \in \{1,..,N_R\}$, at each temperature T_i, with $i \in \{1,..,N_T\}$, denoted by $N_{nm,R}(\Delta t, T_i)$. Accordingly, $G_{T_i}(n,\Delta t \mid m,0)$ is the corresponding transition probability. Thus the likelihood of a *T*-trajectory, running at temperature T_i, can be written as

$$\Lambda_{REMD}(T_i) = \prod_{R=1}^{N_R} \prod_{n,m=0}^{N-1} \left[G_{T_i}(n,\Delta t \mid m,0) \right]^{N_{nm,R}(\Delta t,T_i)} = \prod_{n,m=0}^{N-1} \left[G_{T_i}(n,\Delta t \mid m,0) \right]^{N_{nm}^{REMD}(\Delta t,T_i)} , \qquad (2)$$

where $N_{nm}^{REMD}(\Delta t, T_i) = \sum_{R=1}^{N_R} N_{nm,R}(\Delta t, T_i)$, and the likelihood of an *R*-trajectory becomes accordingly

$$\Lambda_{REMD}(R) = \prod_{i=1}^{N_T} \prod_{n,m=0}^{N-1} \left[G_{T_i}(n, \Delta t \mid m, 0) \right]^{N_{nm,R}(\Delta t, T_i)} = \prod_{n,m=0}^{N-1} \left[G_{REMD}(n, \Delta t \mid m, 0) \right]^{N_{nm,R}(\Delta t)}, \quad (3)$$

where $N_{nm,R}(\Delta t) = \sum_{i=1}^{N_T} N_{nm,R}(\Delta t, T_i)$ is the number of *m* to *n* transitions occurring in the continuous *R*-trajectory of replica *R* regardless of temperature, and

$$G_{REMD}(n, \Delta t \mid m, 0) = \exp\left(\frac{1}{N_{nm,R}(\Delta t)} \sum_{i=1}^{N_T} N_{nm,R}(\Delta t, T_i) \ln G_{T_i}(n, \Delta t \mid m, 0) \right) \quad (4)$$

is the corresponding propagator for the *R*-trajectories.

For converged REMD simulations, all *R*-trajectories are evolving in the same ensemble corresponding to a representative, "intermediate" (i.e., over the entire set of REMD temperatures),[16, 35] dynamics of system replicas. According to Eq. 4, the corresponding replica propagator, $G_{REMD}(n, \Delta t \mid m, 0)$, is effectively the weighted geometric mean of the propagators corresponding to each REMD temperature, T_i.

To further improve statistical estimates, we can combine the propagators extracted from all the *R*-trajectories to construct a single REMD likelihood (i.e., accounting for transitions occurring in all the replicas) written as

$$\Lambda_{REMD} = \prod_{n,m=0}^{N-1} \prod_{R=1}^{N_R} \left[G_{REMD}(n, \Delta t \mid m, 0) \right]^{N_{nm,R}(\Delta t)} = \prod_{n,m=0}^{N-1} \left[G_{REMD}(n, \Delta t \mid m, 0) \right]^{N_{nm}^{REMD}(\Delta t)}, \quad (5)$$

where $N_{nm}^{REMD}(\Delta t) = \sum_{R=1}^{N_R} N_{nm,R}(\Delta t)$ is simply the number of *m* to *n* transitions occurring in all the continuous *R*-trajectories and, typically, $N_R = N_T$.

Importantly, each of the likelihood functions defined above depends on the transition rates of the corresponding system[8] through the relation $G(n,\Delta t \mid m,0)=\left[e^{\mathbf{K}\Delta t}\right]_{nm}$ where **K** is the corresponding *N*-dimensional rate matrix. Note that the optimal elements of the rate matrix **K** can be defined as the rates k_{nm} that maximize the likelihood Λ that a stochastic trajectory corresponding to **K** would have the same number of transitions as collected in the transition matrix **N.** Thus, for both *T*- and *R*-trajectories we can search in the space of possible rates k_{nm} for the ones that maximize each likelihood function in Eqs. 2-4. Note that this is often a non-trivial multidimensional search that can be, however, simplified because the upper- and lower-diagonal elements of the rate matrix are related through detailed balance. Specifically, we perform simulated annealing using a Metropolis Monte Carlo algorithm with rates k_{nm} as parameters, as described before for the analysis of folding rates of monomeric Ala2[34] and Ala5[8-9] molecules. Here, we have thus an initial system with 7 candidate states (Figs. 1(b) and 1(c)). The corresponding 7x7 rate matrices, extracted from the REMD runs at each temperature, are analyzed and shown to converge after using as little as 50% of the total data (i.e., 400 ns of REMD, Figs. S2 and S3). To discuss the dynamics captured by each rate matrix **K**, we compare the relative values of the populations estimated for each state, lifetimes and relaxation times. The convergence of the extracted populations, lifetimes and relaxation times is illustrated in Figs. S2 and S3 for the lowest (310 K) and highest (369.08 K) temperatures, respectively.

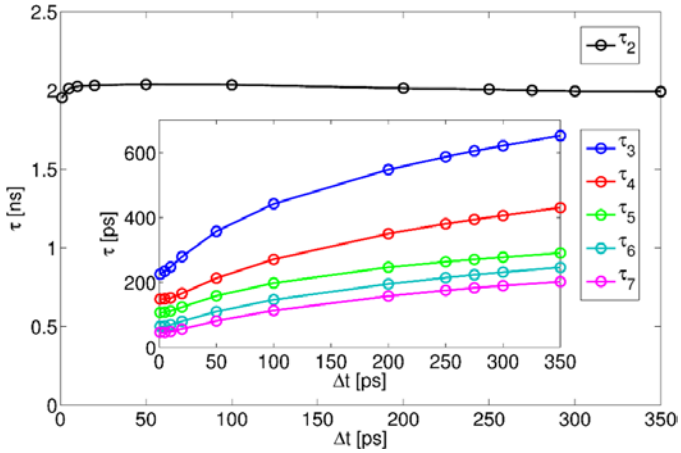


Figure 2. Slowest relaxation time τ_2, related to the overall binding-unbinding process, becomes invariant with propagator lag time. Inset: Lag-time dependence of the faster relaxation times τ_3 to τ_7 of the conformational states of the dimer, estimated from all *R*-trajectories.

As proposed previously, we also study the dependence of our results on the quality of the TBA state assignment used (Fig. S4).[2, 8] As an additional test, we also extract the corresponding dynamics using the *R*-trajectories (Eqs. 3 and 4), with much better statistical data since all replicas are equivalent. We show that the "replica" conformational dynamics is intermediate to the ones corresponding to our lowest and highest REMD temperatures (Fig. S5).[16, 35]

Importantly, when using Markovian transition probabilities to define trajectory likelihoods[8] such as in Eqs. 2-5, monitoring the dependence of the extracted rate matrices on the lag time Δt is a good indication that more complex, sampling-related or possibly non-Markovian kinetic effects could affect the relaxation processes at these scales. Here, we use the continuous *R*-trajectories (often ignored in most REMD simulations) to monitor the dependence of extracted relaxation times on the lag time Δt. This dependence shows clearly (Fig. 2) that, for the NNQQ dimer system, it is only the slowest relaxation time (τ_2, see Fig. 2) that is not Δt-dependent to a very good approximation. For Δton the order of 10 ps or less, non-Markovian effects may influence all propagators extracted from atomistic MD trajectories. In this case, all the fast

relaxation times (Fig. 2 inset) depend to the lag-time used in analysis to a various extent, increasing monotonically with Δt. Together with the observed significant splitting in the eigenvalue spectrum ($\lambda_3/\lambda_2 = \tau_2/\tau_3$) varying between approx. 5 at 310 K and 10 at 369 K, Figs. S2 and S3), this leads us to infer that the kinetics is two state-like to a good approximation,[36-37] most likely due to the significant free energy barrier for the binding/unbinding dynamics of the NNQQ peptides. However, most likely due to the absence of strong barriers between the binding modes S_1 to S_6 (Figs. 1b and 1c) leads to more complex dynamic transitions between them from a Markovian point of view.

To study this further, we project the REMD dynamics on the 2-state bound and dissociated dimer states as illustrated by using distances d_2, d_3 and d_4 in Fig. 3a. Even in this higher-dimensional RC space, it is evidently difficult to separate the truly bound dimers from short-lived "encounter" conformations that are due to non-reactive collisions.

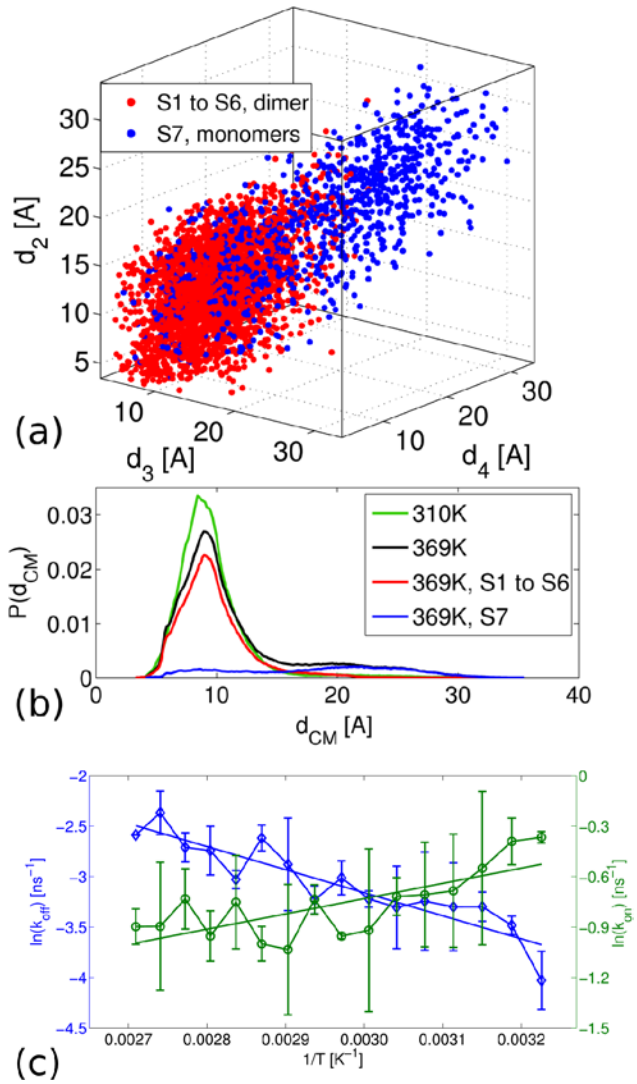


Figure 3. (a) Three dimensional (d_2 vs. d_3 vs. d_4) plot of the distribution of dimer states (red) and dissociated monomer states (blue). (b) The probability distribution of the system as a function of the distance between the centre of mass of each monomer d_{CM}. This allows the deconvolution of $P(d_{CM})$ at 369 K (black) into the dimer contribution (red) and the dissociated monomer states contribution (blue). $P(d_{CM})$ at 310 K (green) is shown for reference. (c) Arrhenius plot of log(k_{off}) (blue diamonds) and log(k_{on}) (green circles) versus 1/T. $k_{on/off}$ were estimated by projecting the 7-state dynamics on a 2-state (binding/unibing) model with the same dissociated state S_7.

This separation (Fig. 3) is enabled by our kinetic analysis and transition-based assignment (TBA) method. In Fig. 3b, we show the probability density along a more typical RC: the center of mass distance d_{CM} for the two monomers. Note that we can thus separate the contribution of

the kinetically dissociated S_7 state from the cumulative distribution of the bound states, S_1 to S_6, and that the probability to be dissociated is higher at larger temperatures. Without any *a priori* assumptions about the functional (e.g., Arrhenius or not) temperature dependence of the rates, we can thus calculate the corresponding binding (k_{on}, Fig. 3c green) and dissociation (k_{off}, Fig. 3c blue) rates, with errors estimated by block averaging.

Also, for efficiency, we simulate a situation with relatively high binding probability, and thus the population of the dissociated S7 state is small. However, these results can be used to estimate the corresponding behavior at different concentrations (i.e., simulation box sizes) by considering that k_{off} would not be expected to depend on concentration, and that the binding process (k_{on}) is diffusion-controlled. We also estimate the corresponding k_{on} and k_{off} rates from *R*-trajectories, both for individual replicas (Fig. S6, blue) and for all the combined *R*-trajectories (Fig. S6, green value) and we show that they are intermediate to the values obtained for *T*-trajectories (Fig. S6, red).

To test convergence, we also show that the fraction of the total REMD simulation time spent by each of the 16 replicas at each temperature is the same to a very good approximation (blue, lines, Fig. S7(a)). Values corresponding to our first initial condition 1B (Fig. 1) are illustrated in Fig. S7. The "equal occupancy" feature of parallel tempering[38] is independent of the choice of ensemble temperatures, and is a useful method for assessing the performance of parallel tempering simulations.[38-39] Additionally, the moments of temperature distributions corresponding to each replica trajectory are essentially the same, as shown in Fig. S7(b).

The *T*-dependent kinetic networks of the NNQQ dimer are illustrated in Fig. 4. We calculate transition fluxes (no. of transitions per ns) between the conformational states of the dimer at T_0 in Fig. 4a and at T_{15} in 4(b). The main (i.e., bigger than 0.01 transitions per 1 ns for transitions

involving S_7 and 0.1 per ns otherwise) fluxes between bounded conformational basins (black), and dissociation/binding rates (red, Markovian transitions) are shown. Note that the backward and forward fluxes should be ideally equal to each other at equilibrium, but they are likely different for actual simulations[8] due to statistical sampling reasons as illustrated in Figs. 4a and 4b. The dimer populations (black) are normalized to the total population of bound states, to emphasize the *T*-dependence of the binding modes. The population of the dissociated state (S_7, red) corresponds to the coarse grained Markovian two-state system. S_2 is the most populated bounded basin (23.8 %) at T_0 and, at all temperatures, the dissociation pathway predominantly occurs through state S_6. The fluxes between S_1 and S_2 are high at T_0 but decrease at T_{15}. The number of transitions between S_1 and S_2 is reduced at higher temperatures, the system spending also less time in either state ($P_{S1}(T_0)$ = 17.8%, $P_{S1}(T_{15})$ = 14.7% and $P_{S2}(T_0)$ = 23.8%, $P_{S2}(T_{15})$ = 19.4%). A similar situation happens between S_2 and S_3 but it is not as pronounced. Nearly all other fluxes increase with temperature, except S_2 to S_7, which, interestingly, flips direction at higher temperatures.

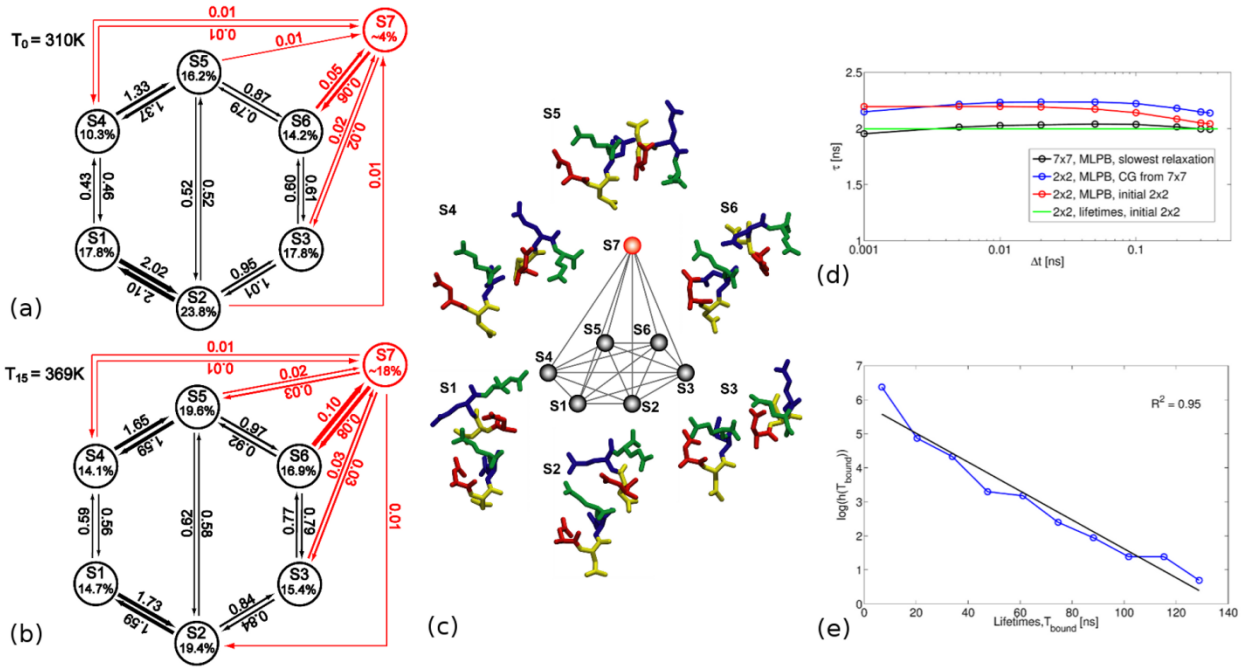


Figure 4. (a) Transition fluxes between bound states of the dimer in black with clearly Markovian fluxes of the binding-unbinding process in red for T_0 (310 K) and for (b) T_{15} (369 K). Only the dominant fluxes are shown with units of ns^{-1}. The arrow widths reflect the size of flux. (c) Representative structures of each of the conformational states of the dimer. (d) Four methods of calculating the slowest relaxation time of the system, at T_{15} (369 K), as a function of the lag time Δt. (e) Log of the distribution of the lifetimes of the dimer state showing single exponential behavior.

Representative structures of the NNQQ dimer states are shown in Fig. 4c with the first N residue in red and the final Q residue in green. Our analysis allows the detailed characterization of these binding modes without the need of additional clustering. Note that, in spite of their small lengths, when part of the S_1-S_6 dimer basins, the NNQQ monomers adopt preferentially non-extended, hairpin-like conformations.[13] However, the inter-basin free energy barriers are all relatively small as compared to the dissociation/binding process, and microcrystal-like[11] extended conformations are also observed with significant probability (see d_{EE} histogram in Fig. S1). As shown in Fig. 4, states S_1 and S_2 (and S_4 and S_5, respectively) interconvert more rapidly than the other states, at all temperatures, and may thus be merged in a more detailed analysis. The emerging binding-unbinding mechanism appears to be dominated at all T by fast inter-conversion between binding modes, with (an order of magnitude slower) dissociation events that occur preferentially from S_6.

While these mechanistic details are expected to be, of course, dependent on choices such as force field, temperature range and other modeling and order parameters used, nevertheless, this study illustrates the broad range of detailed inferences that are enabled by our approach.

As an additional check, the slowest relaxation time, τ_2, of the system is calculated using four different methods as a function of the lag time Δt (Fig. 4d). The value of τ_2 extracted from the initial 7x7 rate matrix calculated by the MLPB method shown is in black. This rate matrix is coarse grained to a 2x2 rate matrix and this value is shown in blue. Another value is obtained by

initially assigning only two states (red), using a radius of 2.5 Å for a circle centered on the dimer basin in Figs. 1b and 1c, and using the same cut-off definition of the dissociated monomers state. The final value of τ_2 is obtained using a lifetime-based method with the data that was initially assigned as a two state system.[8]

Finally, having access to the lifetimes of the system in R-trajectories allows us to monitor their distribution for conformations in the bound dimer basin (Fig. 4e). A single exponential decay is observed, implying that the dissociation of the dimer state is indeed Markovian and that the binding-unbinding process is a two-state kinetic process to a very good approximation.[36-37, 40] Note that the many bound states could correspond to polymorphic structures. Polymorphs of NNQQ are observed to form in microcrystals and this supports the notion that short amyloid protein segments are prone to kinetically driven aggregation.[41-43]

In summary, we show that our CME-based formalism - using transition likelihood maximization and TBA assignment - can be coupled with enhanced sampling REMD simulations with explicit water molecules to characterize the kinetics of dimer formation of short NNQQ amyloid-forming peptides. Importantly, the extracted temperature-dependent kinetic mechanism does not rely on any assumption regarding the functional form (e.g., Arrhenius or not) of the transition rates. Here, we evaluate the rates for both continuous replica trajectories that visit different temperatures, and for discontinuous data corresponding to each REMD temperature. By exhaustively sampling (see Figs. S2 and S3) the conformational states of the system using atomistic REMD, we show that our systematic analysis allows the identification and characterization of fast interconverting binding modes, as well as of the slower binding-unbinding kinetics. Central to our approach is the use of the TBA method that allows us to

control systematically the negative effects of Kramers-type recrossings by assigning conformational states based on their sequential time evolution.[2, 8]

With carefully designed tests (e.g., by monitoring the lag time-dependence of relaxation times, as in Fig. 2, or the exponential decay of state lifetimes, as in Fig. 4e), we probe and quantify the limitations of the *a priori* Markovian assumption on the nature of proposed conformational basins. While the binding-unbinding kinetic process in this case appears to be clearly Markovian, the conformational dynamics of the bound NNQQ dimer has a more complex character. We note that in spite of using several relevant distances as order parameters for extracting the dimerization kinetics, the true character of the transitions between bound states may only be revealed by an additional analysis focused on this specific aspect, and using better reaction coordinates. Nevertheless, our kinetic analysis allows us a quantitative discrimination between short-lived encounter pairs of peptides and stronger bound conformational dimer states (Figs. 3 and 4). While understanding the detailed mechanisms of amyloid peptides interactions has a general biomedical importance, our focus on NNQQ peptides is motivated more by their small size and high biophysical relevance, being some of the first amyloid microcrystal-forming peptides.[11] The conformational dynamics of NNQQ dimers extracted from our REMD simulations supports a kinetically driven aggregation mechanism,[44] in agreement with the polymorphic character reported for amyloid aggregates such as microcrystals and fibrils.[11-12, 45]

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org/journal/jpclcd

ABBREVIATIONS

REMD = replica-exchange molecular dynamics; MD = molecular dynamics, CME = coarse master equation

ACKNOWLEDGMENTS

We thank Drs. Alexander Berezhkovskii, Robert Best, Bernard Brooks, and Attila Szabo from the National Institutes of Health for helpful and stimulating discussions, and the Biowulf cluster at the National Institutes of Health, USA (http://biowulf.nih.gov) for the provision of computational facilities and support. We gratefully acknowledge financial support from the Irish Research Council (IRC) for C.T.L, R.M, and N.V.B., and from the Engineering and Physical Sciences Research Council (EPSRC, grant number EP/N020669/1) for E.R.

REFERENCES

1. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E., Protein Folding Kinetics and Thermodynamics from Atomistic Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17845-50.
2. Buchner, G. S.; Murphy, R. D.; Buchete, N. V.; Kubelka, J., Dynamics of Protein Folding: Probing the Kinetic Network of Folding-Unfolding Transitions with Experiment and Theory. *Biochim. Biophys. Acta, Proteins Proteomics* **2011**, *1814*, 1001-1020.
3. Buchete, N. V., Unlocking the Atomic-Level Details of Amyloid Fibril Growth through Advanced Biomolecular Simulations. *Biophys. J.* **2012**, *103*, 1411-3.
4. Dill, K. A.; MacCallum, J. L., The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042-1046.
5. Best, R. B., Atomistic Molecular Simulations of Protein Folding. *Curr. Opin. Struct. Biol.* **2012**, *22*, 52-61.
6. Hummer, G., Molecular Binding: Under Water's Influence. *Nat. Chem.* **2010**, *2*, 906-907.
7. Sirur, A.; De Sancho, D.; Best, R. B., Markov State Models of Protein Misfolding. *J. Chem. Phys.* **2016**, *144*, 075101.
8. Buchete, N. V.; Hummer, G., Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057-6069.
9. Buchete, N. V.; Hummer, G., Peptide Folding Kinetics from Replica Exchange Molecular Dynamics. *Phys. Rev. E* **2008**, *77*, 4.
10. Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D., Structure of the Cross-Beta Spine of Amyloid-Like Fibrils. *Nature* **2005**, *435*, 773-778.
11. Sawaya, M. R., et al., Atomic Structures of Amyloid Cross-Beta Spines Reveal Varied Steric Zippers. *Nature* **2007**, *447*, 453-457.
12. Wiltzius, J. J. W.; Landau, M.; Nelson, R.; Sawaya, M. R.; Apostol, M. I.; Goldschmidt, L.; Soriaga, A. B.; Cascio, D.; Rajashankar, K.; Eisenberg, D., Molecular Mechanisms for Protein-Encoded Inheritance. *Nat. Struct. Mol. Biol.* **2009**, *16*, 973-U98.
13. Nam, H. B.; Kouza, M.; Hoang, Z.; Li, M. S., Relationship between Population of the Fibril-Prone Conformation in the Monomeric State and Oligomer Formation Times of Peptides: Insights from All-Atom Simulations. *J. Chem. Phys.* **2010**, *132*, 10.
14. Lu, Y.; Wei, G. H.; Derreumaux, P., Structural, Thermodynamical, and Dynamical Properties of Oligomers Formed by the Amyloid Nnqq Peptide: Insights from Coarse-Grained Simulations. *J. Chem. Phys.* **2012**, *137*, 8.

15. Sugita, Y.; Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141-151.
16. Rosta, E.; Hummer, G., Error and Efficiency of Replica Exchange Molecular Dynamics Simulations. *J. Chem. Phys.* **2009**, *131*, 12.
17. Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M., Recovering Kinetics from a Simplified Protein Folding Model Using Replica Exchange Simulations: A Kinetic Network and Effective Stochastic Dynamics. *J. Phys. Chem. B* **2009**, *113*, 11702-11709.
18. Xia, J.; Flynn, W. F.; Gallicchio, E.; Zhang, B. W.; He, P.; Tan, Z.; Levy, R. M., Large-Scale Asynchronous and Distributed Multidimensional Replica Exchange Molecular Simulations and Efficiency Analysis. *Journal of Computational Chemistry* **2015**, *36*, 1772-1785.
19. Murphy, R. D.; Conlon, J.; Mansoor, T.; Luca, S.; Vaiana, S. M.; Buchete, N.-V., Conformational Dynamics of Human Iapp Monomers. *Biophys. Chem.* **2012**, *167*, 1-7.
20. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F., Pyemma 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525-5542.
21. Plattner, N.; Noe, F., Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat. Commun.* **2015**, *6*.
22. Duan, M.; Liu, H.; Li, M.; Huo, S., Network Representation of Conformational Transitions between Hidden Intermediates of Rd-Apocytochrome B562. *J. Chem. Phys.* **2015**, *143*, 135101.
23. Peters, B.; Bolhuis, P. G.; Mullen, R. G.; Shea, J.-E., Reaction Coordinates, One-Dimensional Smoluchowski Equations, and a Test for Dynamical Self-Consistency. *J. Chem. Phys.* **2013**, *138*, 054106.
24. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F., Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139*, 015102.
25. Schwantes, C. R.; Pande, V. S., Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of Ntl9. *J. Chem. Theory Comput.* **2013**, *9*, 2000-2009.
26. Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.
27. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
28. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712-725.
29. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.
30. Patriksson, A.; van der Spoel, D., A Temperature Predictor for Parallel Tempering Simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073-2077.
31. Pastor, R. W.; Brooks, B. R.; Szabo, A., An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65*, 1409-1419.

32. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
33. Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E., Exchange Often and Properly in Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 2804-2808.
34. Sriraman, S.; Kevrekidis, I. G.; Hummer, G., Coarse Master Equation from Bayesian Analysis of Replica Molecular Dynamics Simulations. *J. Phys. Chem. B* **2005**, *109*, 6479-84.
35. Rosta, E.; Hummer, G., Error and Efficiency of Simulated Tempering Simulations. *J. Chem. Phys.* **2010**, *132*, 034102.
36. Berezhkovskii, A. M.; Murphy, R. D.; Buchete, N.-V., Note: Network Random Walk Model of Two-State Protein Folding: Test of the Theory. *J. Chem. Phys.* **2013**, *138*, 036101-2.
37. Berezhkovskii, A. M.; Tofoleanu, F.; Buchete, N. V., Are Peptides Good Two-State Folders? *J. Chem. Theory Comput.* **2011**, *7*, 2370-2375.
38. Doll, J. D.; Dupuis, P., On Performance Measures for Infinite Swapping Monte Carlo Methods. *The Journal of Chemical Physics* **2015**, *142*, 024111.
39. Doll, J. D.; Plattner, N.; Freeman, D. L.; Liu, Y.; Dupuis, P., Rare-Event Sampling: Occupation-Based Performance Measures for Parallel Tempering and Infinite Swapping Monte Carlo Methods. *The Journal of Chemical Physics* **2012**, *137*, 204112.
40. Hummer, G.; Szabo, A., Optimal Dimensionality Reduction of Multistate Kinetic and Markov-State Models. *J. Phys. Chem. B* **2015**, *119*, 9029-9037.
41. Park, J.; Kahng, B.; Hwang, W., Thermodynamic Selection of Steric Zipper Patterns in the Amyloid Cross-Beta Spine. *PLoS Comput. Biol.* **2009**, *5*, 17.
42. Pellarin, R.; Schuetz, P.; Guarnera, E.; Caflisch, A., Amyloid Fibril Polymorphism Is under Kinetic Control. *J. Am. Chem. Soc.* **2010**, *132*, 14960-14970.
43. Morriss-Andrews, A.; Shea, J.-E., Simulations of Protein Aggregation: Insights from Atomistic and Coarse-Grained Models. *J. Phys. Chem. Lett.* **2014**, *5*, 1899-1908.
44. Hwang, W.; Zhang, S. G.; Kamm, R. D.; Karplus, M., Kinetic Control of Dimer Structure Formation in Amyloid Fibrillogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12916-12921.
45. Berhanu, W. M.; Masunov, A. E., Alternative Packing Modes Leading to Amyloid Polymorphism in Five Fragments Studied with Molecular Dynamics. *Biopolymers* **2012**, *98*, 131-44.